Field Crops Research 224 (2018) 37–47

Contents lists available at ScienceDirect

Field Crops Research

journal homepage: www.elsevier.com/locate/fcr

Modelling grass yields in northern climates – a comparison of three growth models for timothy

Panu Korhonen[a,*], Taru Palosuo[a], Tomas Persson[b], Mats Höglind[b], Guillaume Jégo[c], Marcel Van Oijen[d], Anne-Maj Gustavsson[e], Gilles Bélanger[c], Perttu Virkajärvi[a]

[a] Natural Resources Institute Finland (Luke), Halolantie 31A, 71750, Kuopio, Finland
[b] Norwegian Institute of Bioeconomy Research (NIBIO), Postvegen 213, 4353, Klepp Stasjon, Norway
[c] Agriculture and Agri-Food Canada (AAFC), 2560 Hochelaga Boulevard, Québec City, Québec, G1 V 2J3, Canada
[d] Centre for Ecology and Hydrology (CEH), Bush Estate, Penicuik, EH26 0QB, UK
[e] Swedish University of Agricultural Sciences (SLU), SE-901 83, Umeå, Sweden

ARTICLE INFO

Keywords:
Forage grass
Model comparison
Timothy
Uncertainty
Yield

ABSTRACT

During the past few years, several studies have compared the performance of crop simulation models to assess the uncertainties in model-based climate change impact assessments and other modelling studies. Many of these studies have concentrated on cereal crops, while fewer model comparisons have been conducted for grasses. We compared the predictions for timothy grass (*Phleum pratense* L.) yields for first and second cuts along with the dynamics of above-ground biomass for the grass simulation models BASGRA and CATIMO, and the soil-crop model STICS. The models were calibrated and evaluated using field data from seven sites across Northern Europe and Canada with different climates, soil conditions and management practices. Altogether the models were compared using data on timothy grass from 33 combinations of sites, cultivars and management regimes. Model performances with two calibration approaches, cultivar-specific and generic calibrations, were compared. All the models studied estimated the dynamics of above-ground biomass and the leaf area index satisfactorily, but tended to underestimate the first cut yield. Cultivar-specific calibration resulted in more accurate first cut yield predictions than the generic calibration achieving root mean square errors approximately one third lower for the cultivar-specific calibration. For the second cut, the difference between the calibration methods was small. The results indicate that detailed soil process descriptions improved the overall model performance and the model responses to management, such as nitrogen applications. The results also suggest that taking the genetic variability into account between cultivars of timothy grass also improves the yield estimates. Calibrations using both spring and summer growth data simultaneously revealed that processes determining the growth in these two periods require further attention in model development.

1. Introduction

Process-based crop simulation models that simulate crop growth, development, and yields, while taking into account the interactions between the crop genotype, management and environmental factors are increasingly used to support decision making and planning in agriculture, including aspects related to animal feed and forage production (Kipling et al., 2016). Several studies have recently been published on the comparison of the performance of crop simulation models under different environmental conditions in an effort to improve crop models and climate impact assessment projections and to gain an understanding of the uncertainties related to these assessments (see, e.g. Asseng et al., 2013; Asseng et al., 2015; Bassu et al., 2014; Pirttioja et al., 2015). In addition, there have been model-based evaluations of adaptation options to climate change (Ruiz-Ramos et al., 2017; Chenu et al., 2017).

To date, model comparisons and model ensemble studies have mostly focused on cereal crops and fewer model comparisons have been published for perennial forage grasses. Still, many crop models or crop modules of farm system models, e.g. STICS (Jégo et al., 2013) and APSIM (Keating et al., 2003), can also simulate forage grasses. There are also separate forage grass models (e.g. BASGRA, Höglind et al., 2016; CATIMO, Bonesmo and Bélanger, 2002a) that have comparable process descriptions to those in cereal crop models, such as radiation interception and use efficiency. Forage grass production systems, however, have specific characteristics that should be taken into account

* Corresponding author.
E-mail address: panu.korhonen@luke.fi (P. Korhonen).

https://doi.org/10.1016/j.fcr.2018.04.014
Received 16 June 2017; Received in revised form 27 March 2018; Accepted 26 April 2018
Available online 24 May 2018
0378-4290/ © 2018 Elsevier B.V. All rights reserved.

in simulation models. Almost all above-ground biomass is harvested several times during growing seasons and consequently the status of the plants after cuttings and regrowth are key issues for forage grass models (Jing et al., 2013). Another important aspect is the dynamically changing feed quality during forage development (Bonesmo and Bélanger, 2002b; Gustavsson and Martinsson, 2001). Finally, grass leys are typically perennial, which makes it essential to simulate the growth initiation in the spring (Bélanger et al., 2008) and thus, makes it important to develop the models to simulate relevant processes related to over-wintering particularly for forage grasses at high latitudes, where there is virtually no cold-season growth (Höglind et al., 2016).

To our knowledge, those forage grass model comparisons that have been conducted (e.g. Hurtado-Uria et al., 2013) were restricted to regions with rather homogeneous climate conditions, without testing their performance and suitability over a wide range of climate conditions. Even though the forage grass modules of current farm system models and the separate forage grass models that are referred to above are considered process-based, they all include several empirically derived functions. Therefore, one could assume that the predictability of such forage grass models or modules would vary with climatic and other environmental conditions. Hence, a comparison of crop simulation models across a wide range of conditions is a key to strengthening the understanding of the effects of different model process descriptions, i.e. model structures, on yield and quality related output variables. Moreover, varying the genetic variability and climate and soil conditions within the calibration and evaluation datasets could provide knowledge about the calibration procedures as well as knowledge about model application strategies. Persson et al. (2014) found that the observed dry matter yield for one variety (cv Grindstad) was more accurately predicted by LINGRA model (the predecessor of the BASGRA model) when the parameters were calibrated against data from several locations within a region with heterogeneous climate and soil conditions than when the parameters were calibrated against data from only one location. However, to the best of our knowledge no study has been published where the effects varying the genetic variability in calibration datasets on grassland dry matter yield were evaluated. Such a study providing knowledge about model sensitivity to genetic variability could be used to arrange field trial data for model calibration. It could also give useful information about how to calibrate and apply grassland models for genetically heterogeneous conditions, such as in estimations of regional or national grassland productivity.

Timothy (*Phleum pratense* L.) is one of the most important forage grass species in the cold temperate climate zone of the northern hemisphere, including Canada and the Nordic countries in Europe. Management of timothy swards varies considerably according to the climate and soil conditions where it is grown and with its end use. Timothy is grown either in pure stands or mixed with other grasses or leguminous species for three years or longer before it is ploughed up and reseeded or rotated with an annual crop. When used to feed dairy cows, timothy is often cut and harvested at the mid-heading stage to optimize the nutritive value and yield. When used as feed for beef cattle and sheep, timothy is usually cut at later stages, which generally results in higher yields but a lower nutritive value. The number of harvests per year usually varies from two to four depending on the cutting strategy, the cultivar-specific characteristics such as the development rate and its effect on nutrient composition, as well as the climate and weather conditions. In addition, plant characteristics, such as the maximum tiller height, pattern of development of vegetative and reproductive tillers and the leaf/stem ratio, vary between cultivars, which have been bred to meet different regional climate conditions and management practices (Virkajärvi et al., 2010). Considering the range of environmental conditions, alternative management strategies and the genetic variability of timothy, a model comparison with timothy data covering different environmental conditions, a wide set of different cultivars and alternative management options would provide material for critical testing of crop simulation models.

The overall aim of this study was to assess and compare the ability of simulation models to accurately simulate the growth and yield of the first and second annual cuts of timothy under different environmental conditions. To this end, the performance of two grass simulation models, BASGRA and CATIMO and the soil-crop model STICS were assessed with a comprehensive experimental dataset collected from across Northern Europe and Canada with varying management practices. The three models were calibrated either specifically for each cultivar (cultivar-specific calibration) or for a number of cultivars all together (generic calibrations) and the performance of the models with both calibrations was tested.

2. Materials and methods

2.1. Models

The three models simulating the growth and development of timothy as a function of weather, soil and crop management factors included in the comparison were: CATIMO (R-version 1.0; Bonesmo and Bélanger, 2002a,b; Jing et al., 2012, 2013), BASGRA (version 2014; Höglind et al., 2001; van Oijen et al., 2005; Höglind et al., 2016) and STICS (v8.4; Brisson et al., 1998, 2008; Jégo et al., 2013; Jing et al., 2017). All three models use the radiation use efficiency (RUE) approach instead of calculating the photosynthesis and respiration in detail and they use a simple leaf area index (LAI) to calculate light interception (Table 1). They all cover soil water and N effects in the simulation of

Table 1
Approaches used by the three models for the major processes determining crop growth and development.

Process	BASGRA	CATIMO	STICS
Leaf area development and light interception[a]	S	S	S
Light utilization[b]	RUE	RUE	RUE
Root distribution over depth[c]	–	–	Sig
Drought stress[d]	ETa/ETp	ETa/ETp	ETa/ETp
Water dynamics[e]	C	C	C
Evapo-transpiration[f]	PM	PM	P
Effect of nitrogen[g]	NSNS	RNC	RNC
Tillering dynamics[h]	C	–	–
Vernalisation[i]	SV	–	–
Start of spring growth[j]	5D	5Drm	1D
Regrowth dynamics[k]	SSDG	RDG	RDG
Soil C/N model[l]	CN, P(3)	N	CN, P(3), B

[a] Leaf area development and light interception: S = simple approach (e.g. leaf area index (LAI).

[b] Light utilization or biomass growth: RUE = simple (descriptive) radiation use efficiency approach.

[c] Root distribution over depth: Sig = sigmoidal.

[d] Drought stress: ETa/ETp = actual to potential evapotranspiration ratio.

[e] Water dynamics approach: C = capacity approach.

[f] Method to calculate evapotranspiration: P = Penman; PM = Penman–Monteith.

[g] Effect of nitrogen: RNC = relative nitrogen concentration as the ratio of the actual N concentration to the critical N concentration (Bélanger and Richards, 1997), NSNS = N source/N sink balance dependent growth.

[h] Tillering dynamics: C = Three different tiller categories (dependent on internal as well as external factors).

[i] Vernalisation: SV = simple approach (threshold temperature).

[j] 5D = the first day of 5 consecutive days above base temperature, 5Drm = the first day when the running mean of a five-day daily mean temperature is above base temperature; 1D = 1 day above base temperature (start defined by model user).

[k] Regrowth dynamics: RDG = reserve dependent growth, SSDG = source (LAI, reserves) and sink (tillers) dependent regrowth.

[l] Soil C/N model: CN = soil CN model, N = soil N model with only mineral N; P(x) = number of organic matter pools; B = microbial biomass pool.

grass growth, but the effect of excess water is only taken into account by the STICS model. The representation of soil processes differs considerably between the three models. While in the STICS model there are detailed sub-routines for soil water and N balances, the CATIMO and BASGRA models are simpler with a description of the root zone as one single homogenous layer. Tillering dynamics and vernalisation are simulated only in BASGRA. The CATIMO model starts biomass accumulation in spring when the daily mean temperature exceeds the base temperature (default 5 °C) (Table 1). In this study, over-wintering was not included and we initiated the spring growth in the BASGRA simulation the first time of the year that the mean air temperature exceeded 5 °C for five consecutive days or longer. In the STICS simulation, biomass accumulation starts after model initialization as soon as the daily average air temperature is above the minimum threshold temperature for net photosynthesis (parameter *temin*). In the case of multi-annual simulations, winter dormancy processes can also be simulated with STICS, but they were not used in this study.

2.2. Experimental sites

For the model intercomparison, crop data were obtained from field experiments at sites across Northern Europe and Canada with different cultivars (Table 2, Fig. 1).

Together, these sites comprise a wide range of climate, soil and management practices that are associated with timothy production in its main production regions of Canada and Northern Europe. Data from two to three growing seasons per site were available and they covered altogether 33 different combinations of site, year, cultivar and management practices (below called treatments) (Table S1). The cultivars differed between the sites, except that the cultivar 'Champ' was used in both Fredericton and Québec (Table 2).

2.3. Setup of model comparison

2.3.1. Information provided for model calibration and runs

Comprehensive data on crop performance and management, as well as soil and weather conditions were obtained from all sites in the study to allow for detailed model calibrations and evaluations. These were conducted for each model independently. We randomly selected one treatment for model evaluation from each treatment type from sites with more than two treatments or years, i.e. Rovaniemi, Særheim, Québec, and Fredericton (in total 9 treatments) and used the remaining 24 for model calibration (Table S1). This approach differs from previous model comparisons (Palosuo et al., 2011; Rötter et al., 2012; Asseng et al., 2013) that applied a "blind test" approach, i.e. model users were not provided with the observed data for crop yields, biomasses or other variables delivered as model results. Here instead, we tried to comprehensively apply all the data available for calibration treatments.

The input data included observed daily weather (i.e. minimum and maximum temperature, precipitation, relative air humidity, wind speed and global radiation), sward and soil management (cut height, cutting and fertilisation dates and amounts) and basic information on soil properties (soil texture, hydraulic characteristics and the pH). Standard deviations (SD) were calculated for the Maaninka and Rovaniemi crop datasets based on original replicates. For other datasets, replicate data were not available and the SDs were calculated assuming a coefficient of variation with a value of 0.05 for the yield, 0.1 for LAI, and 0.5 for other observations.

2.3.2. Simulation setup

The models were run on an annual basis, starting simulations for each growing season in the spring. The starting dates for the simulations were determined with the typical criteria used with each model. For the CATIMO model the start of simulations was determined by a threshold temperature for the running mean of a five-day daily mean temperature (default 5 °C). For BASGRA, the simulations started the

Table 2
Characteristics of experimental sites and their climatic conditions for the study years.

Location	Position lat/lon/alt (m.a.s.l)	Precipitation (mm yr^{-1})	Mean annual temperature (°C)	Data description	Period	Cultivar	Parallel treatments	Observed crop data[a]
Maaninka, Finland	63.14N/27.32E/90 m	560	4.2	Virkajärvi et al. (2012)	2006–2007	Tammisto II	–	DM, sward height, LAI
Rovaniemi, Finland	66.35N/26.01E/106 m	610	1.0	Nissinen et al. (2010)	1999–2001	Iki	two N application rates	DM, sward height, DM_{leaf}, DM_{stem}
Umeå, Sweden	63.45N/20.17E/12 m	595	3.3	Gustavsson and Martinsson (2001)	1995–1996	Jonatan	–	DM, sward height, DM_{leaf}, DM_{stem}
Særheim, Norway	58.36N/5.39E/90 m	1392	8.0	Höglind et al. (2005)	2000–2002	Grindstad	early and late cutting regimes	DM, LAI, SLA, tiller density
Fredericton, Canada	45.55N/66.32W/35 m	1108	5.7	Bélanger and Richards (1997)	1991–1993	Champ	varying N application rates	DM, DM_{leaf}, DM_{stem}, LAI
Québec, Canada	46.47N/71.07W/75 m	1009	5.3	Bélanger et al. (2008)	1999–2001	Champ	three N application rates	DM
Lacombe, Canada	52.28N/113.44W/860 m	429	3.5	Jing et al. (2012); Jégo et al. (2013)	2004–2005	Climax	–	DM, LAI

[a] Abbreviations: DM = dry matter yield, DM_{leaf} = dry matter yield of leaves, DM_{stem} = dry matter yield of stems, LAI = leaf area index, SLA = specific leaf area.

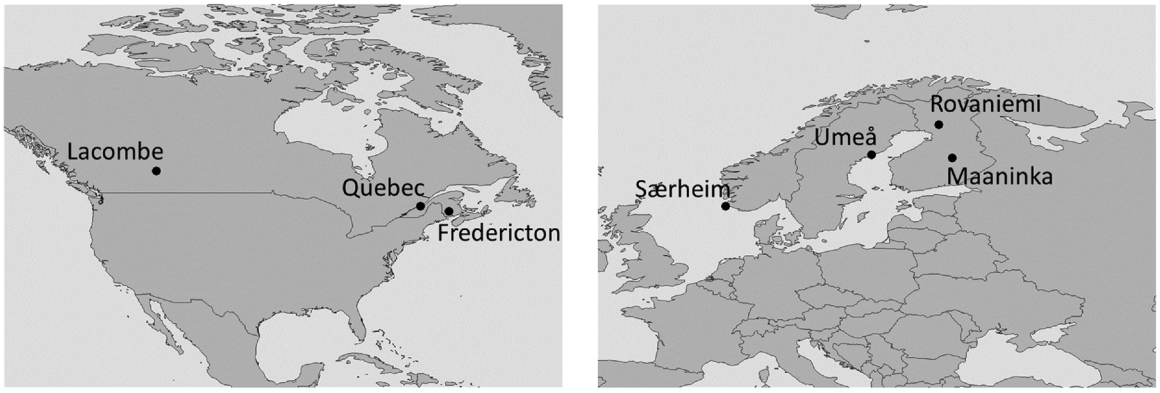

Fig. 1. Locations of the experimental sites.

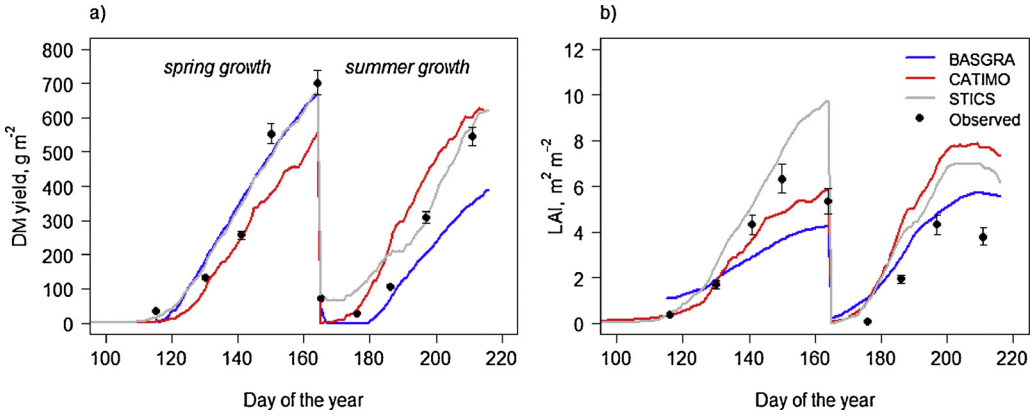

Fig. 2. Simulated (cultivar-specific calibration) and observed (a) above-ground biomass (dry matter) and (b) leaf area index (LAI), using three timothy models at Særheim in 2001 (treatment 11). Error bars represent ± 1 standard deviation. The above-ground biomass dynamics for all sites are shown in Fig. S1 and LAI dynamics in Fig. S2.

first time of the year that there were at least five consecutive days with a mean daily air temperature above 5 °C. For the STICS model, simulations were initialized on 15 April at all sites, except Rovaniemi (10 May) to account for the very short growing season and at Særheim (1 March) to account for the mild winter conditions prevailing on the south-west coast of Norway. In STICS, the initial state variables were set as proposed by Jégo et al. (2013): initial LAI and living shoot dry biomass were set to 0, initial reserve biomass was set to 100 g DM m-2 and the initial root depth was set to 0.50 m. Although timothy swards are perennial, winter seasons were not simulated because not all of the compared models covered the relevant winter processes, and the focus in this study was on comparing models for the growth estimates during the growing seasons. Simulations were carried out using a daily time step in all models and they lasted until the last observation of summer growth in the observed data.

2.3.3. Calibration of the models

The models were calibrated by adjusting their parameters so that their output fitted the provided observed data either: 1) separately for each cultivar (*cultivar-specific calibration*), or 2) using data for all cultivars at all sites together (*generic calibration*). The BASGRA and CATIMO models were calibrated using Bayesian techniques. Both calibrations were performed by means of Markov Chain Monte Carlo (MCMC) sampling using the Metropolis algorithm (van Oijen et al., 2005). Beta distributions for all parameters under calibration were used. The chain length, i.e. the number of model runs was 150,000 with BASGRA and varied from 150,000 to 200,000 with CATIMO. The prior ranges and the maximum a posteriori (MAP) values of the parameters for BASGRA and CATIMO are listed in Tables S2 and S3, respectively. 75 for BASGRA and 8 parameters for CATIMO were adjusted for both cultivar-specific and generic calibrations. Parameters that were not included in the calibration and the prior probability distributions of calibrated parameters for CATIMO were taken from earlier CATIMO studies (Bonesmo and Bélanger, 2002a,b; Jing et al., 2011, 2012). The prior probability distributions for BASGRA were based on common knowledge about the parameters and assumptions taking into account results from previous model evaluations for the cultivar Grindstad under conditions representing northern Europe (Persson et al., 2014).

The STICS model calibration was done using the STICS built-in optimization tool based on a simplex algorithm (Nelder and Mead, 1965). For the generic calibration, 17 parameters (Table S4) were simultaneously calibrated by minimizing the RMSE between simulated and measured biomass and the LAI. Values of other parameters were those used in Jégo et al. (2013). For the cultivar-specific calibration, a sub-selection of 6 parameters (identified as “cultivar parameters” in STICS) was calibrated, while for the other 11 parameters the same values as in the generic calibration were used.

2.3.4. Methods used for evaluating model performance

First, the model performances for the timothy growth were evaluated by comparing sequential above-ground biomass observations with the corresponding simulated values during growth (Fig. 2). This was done for spring growth (from the start of growth in spring to the first cut) and summer growth (from the first cut to the second cut) periods separately. Then, the model performance for the first and second cut yields were assessed, i.e. the observed and simulated above-ground biomass for the last observed data point of spring and summer growths were compared.

The models and the calibration methods were evaluated based on their prediction of above-ground biomass and LAI dynamics as well as first and second cut yields. Observed measurements were considered true values, though we acknowledge that there is always uncertainty in the measurements. The root mean square error (RMSE) was used to show the average difference between the model estimates and the measurements.

$$RMSE = \sqrt{N^{-1}\sum_{i=1}^{N}(P_i - O_i)^2} \quad (1)$$

where O_i is the observed value and P_i is the model prediction and N is the number of estimate-observation pairs. In addition, normalized RMSE values (RMSE%) were calculated by dividing RMSEs by the mean values of observations.

The relative mean bias error (rMBE) was used as an indicator to describe whether the model-predictions were over- or underestimates and their relative magnitude.

$$rMBE = \frac{N^{-1}\sum_{i=1}^{N}(P_i - O_i)}{N^{-1}\sum_{i=1}^{N}(O_i)} \quad (2)$$

Simulated yield responses to spring N fertilisation were analysed using correlations between fertiliser application rates and simulated final spring growth yields for all experiments and by comparing these results to corresponding correlations for observed yields.

3. Results

3.1. Dynamics of above-ground biomass and leaf area index

The start of spring growth varied between the models with STICS simulating a 21-day earlier start of spring growth on average than the other two models with the cultivar-specific calibration (Fig. S1). With the cultivar-specific calibration, the spring growth was simulated most accurately by STICS (rMBE = −0.01). BASGRA (rMBE = −0.12) and CATIMO (rMBE = −0.13) tended to underestimate the above-ground biomass (Fig. 3).

The summer growth was best reproduced by the BASGRA model, albeit with an underestimation (rMBE = −0.16), whereas STICS (rMBE = 0.22) and CATIMO (rMBE = 0.27) overestimated the above-ground biomass for first summer growth. In terms of RMSEs, the model performance for the spring growth for the treatments used for model evaluation was best in the STICS model (Table 3). For the summer growth and the resulting second cut, however, the RMSEs in the BASGRA and STICS models were similar but fewer than in the CATIMO model (Table 3).

The accuracy of the simulated above-ground biomass dynamics for calibration treatments was better than that for the treatments used for evaluating model performance; the RMSEs for the spring growth were from 10% to 55% smaller for the calibration treatments than for the evaluation treatments depending on the model (Table 3).

For all three models, the cultivar-specific calibration resulted in greater simulation accuracy than the generic calibration for the spring growth (Fig. 3). The mean RMSEs of all treatments were smaller with the cultivar-specific calibration than with the generic calibration for BASGRA (-33%), CATIMO (-33%) and STICS (-31%) (Table 3). For the summer growth, the mean RMSEs for all treatments were smaller with the cultivar-specific calibration than with the generic calibration for BASGRA (-16%) and STICS (-12%), whereas for CATIMO the mean RMSEs were 34% higher with the cultivar-specific calibration than with the generic calibration (Table 3).

All three models performed relatively well in simulating LAI dynamics (Fig. S2). The RMSEs for calibration (altogether over 12 treatments at four sites) and evaluation (over two treatments in Særheim and two in Fredericton) were 1.66 and 1.75 $m^2\,m^{-2}$ for BASGRA, 1.72 and 1.57 $m^2\,m^{-2}$ for CATIMO and 1.25 and 1.69 $m^2\,m^{-2}$ for STICS, respectively. The CATIMO model tended to simulate high LAI values for the summer growth period when LAI data was not available for calibration (Fig. S2).

3.2. First and second cut yields

For both cultivar-specific and generic calibrations, the first cut yields were mostly underestimated by all models (Fig. 4). The estimates simulated by STICS were closer to the observed yield than those simulated by BASGRA and CATIMO. The first cut performance of all models for the model evaluation treatments was somewhat weaker than their performance for calibration treatments. For the second cut, STICS provided the most accurate yield estimates, while BASGRA somewhat underestimated the yields on average on all sites and CATIMO systematically overestimated the second cut yields in Québec (Fig. 4).

The ranges of the simulated estimates of first and second cut yields between the models were not affected by the calibration method used except for the second cut in Québec where the range clearly widened when the cultivar-specific calibration was used (Fig. 5). The ranges of both calibration methods (generic or cultivar-specific) were close to each other and no systematic differences were observed. In most treatments, the ranges of both calibration methods overlapped, at least partially. However, at some sites, the effect of the calibration methods was more noticeable. For example, in Québec the first cut yield estimates with the generic calibration were clearly lower than the estimates with the cultivar-specific calibration. In Rovaniemi and Lacombe, on the other hand, the generic calibration provided higher first cut yield estimates than the cultivar-specific calibration.

The observed first cut yield was within the simulated range for 18 treatments (Fig. 5). For the 15 remaining treatments, none of the models/calibrations could achieve the observed first cut yields, which indicates systematic under-prediction in the models. For the second cut, most of the observed yields were included in the simulated range except for one treatment in Særheim (model underprediction) and two treatments in Québec (model overprediction).

The yield estimates calculated as means and medians over the three models for first and second cuts using cultivar specific calibration were better in terms of the RMSEs for all sites, than the yield estimates of BASGRA or CATIMO (data not shown). The first cut yield estimates by the STICS model were more accurate or about same performance level than both the multi-model mean and median with both calibration methods.

Overall, the first cut yields correlated positively with nitrogen fertilisation in both the observed and simulated data over all studied sites and treatments (Fig. 6). The STICS model with a more detailed description of soil N mineralisation and water balance than the other two models managed to simulate yield responses to increasing N fertilisation more accurately than the BASGRA and CATIMO models, which both underestimated the N fertilisation response.

4. Discussion

4.1. Model performances

This study is, to our knowledge, the first comparison of yield simulations from models to simulate timothy, and it shows that the three models estimated the dynamics of above-ground biomass and LAI satisfactorily, although the estimated ranges of the simulation results between the models were quite wide. Despite several sequential observations of the above-ground biomass and LAI provided for model calibration, the model results for many treatments did not match the observed values. The simulated values for summer growth and the resulting yield for the second cut were generally very good, although one model (CATIMO) could not cover the obvious N limitation at one site (Québec).

The simulation accuracy of the three models was roughly the same as the model performances documented in earlier studies of these three models (Jing et al., 2012; Jégo et al., 2013; Persson et al., 2014). The RMSE% for the yields of evaluation treatments with cultivar-specific calibration varied between 22% and 45% for the first cut and between

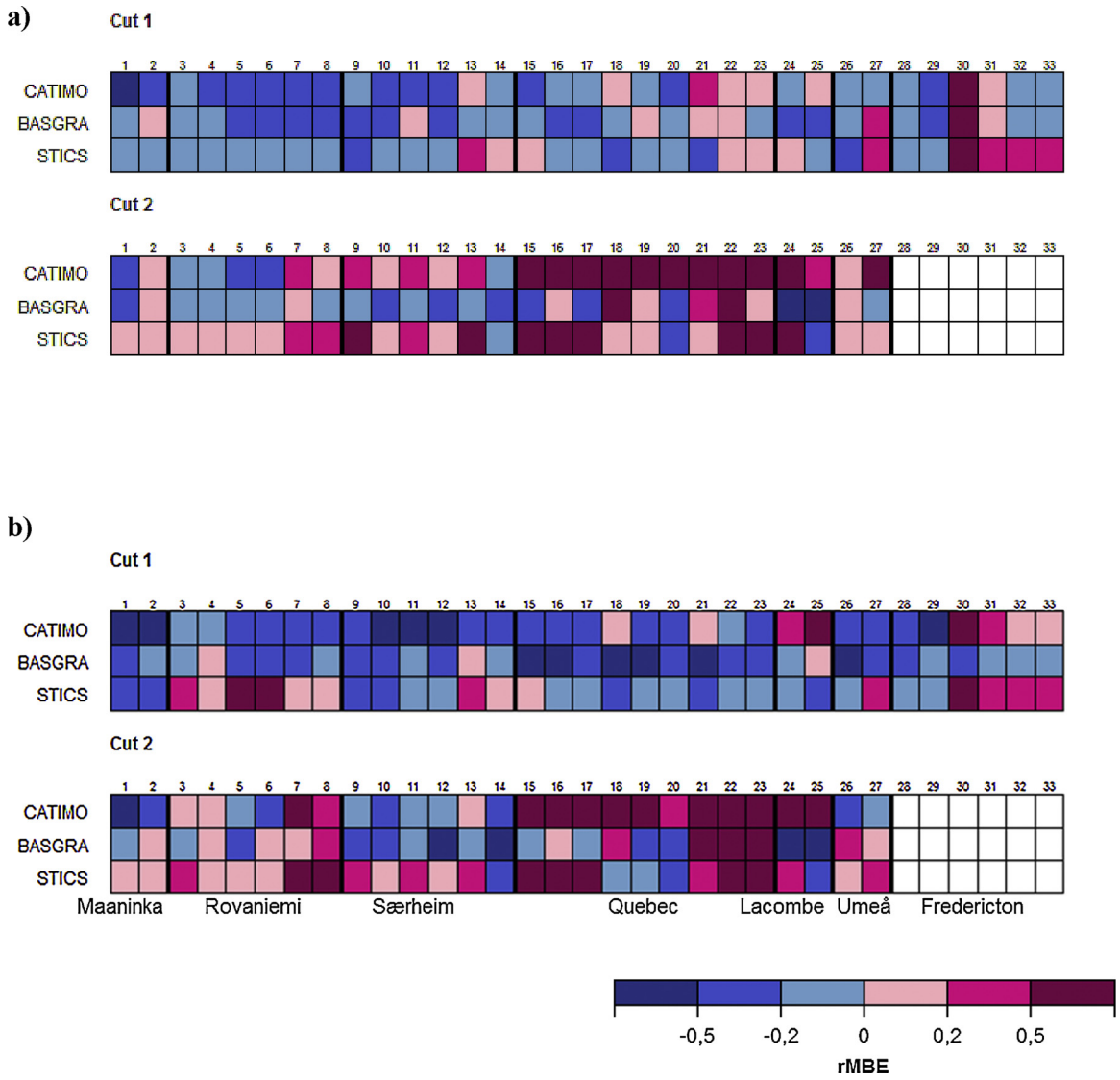

Fig. 3. Relative mean bias error (rMBE) of the simulated above-ground biomass of three timothy models grouped in quantiles with overestimations shown in red shades and underestimations shown in blue shades (see colour scale) during the spring growth (Cut 1) and summer growth (Cut 2) for all 33 treatments simulated with (a) cultivar-specific calibration and (b) generic calibration.

Table 3
The mean of site-specific root mean square errors (RMSE) for the simulated above-ground biomass (g DM m^{-2}) from the calibration and model evaluations of three timothy models during the spring and summer growth periods.

Model	BASGRA				CATIMO				STICS			
Calibration	cultivar-specific		generic		cultivar-specific		generic		cultivar-specific		generic	
	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth	Spring growth	Summer growth
Mean over all treatments (n = 33)	120	73	178	87	113	190	168	142	87	92	126	104
Mean over calibration treatments (n = 24)	90	70	159	86	105	183	167	146	85	95	116	111
Mean over evaluation treatments (n = 9)	199	83	227	93	135	210	171	132	94	81	153	83

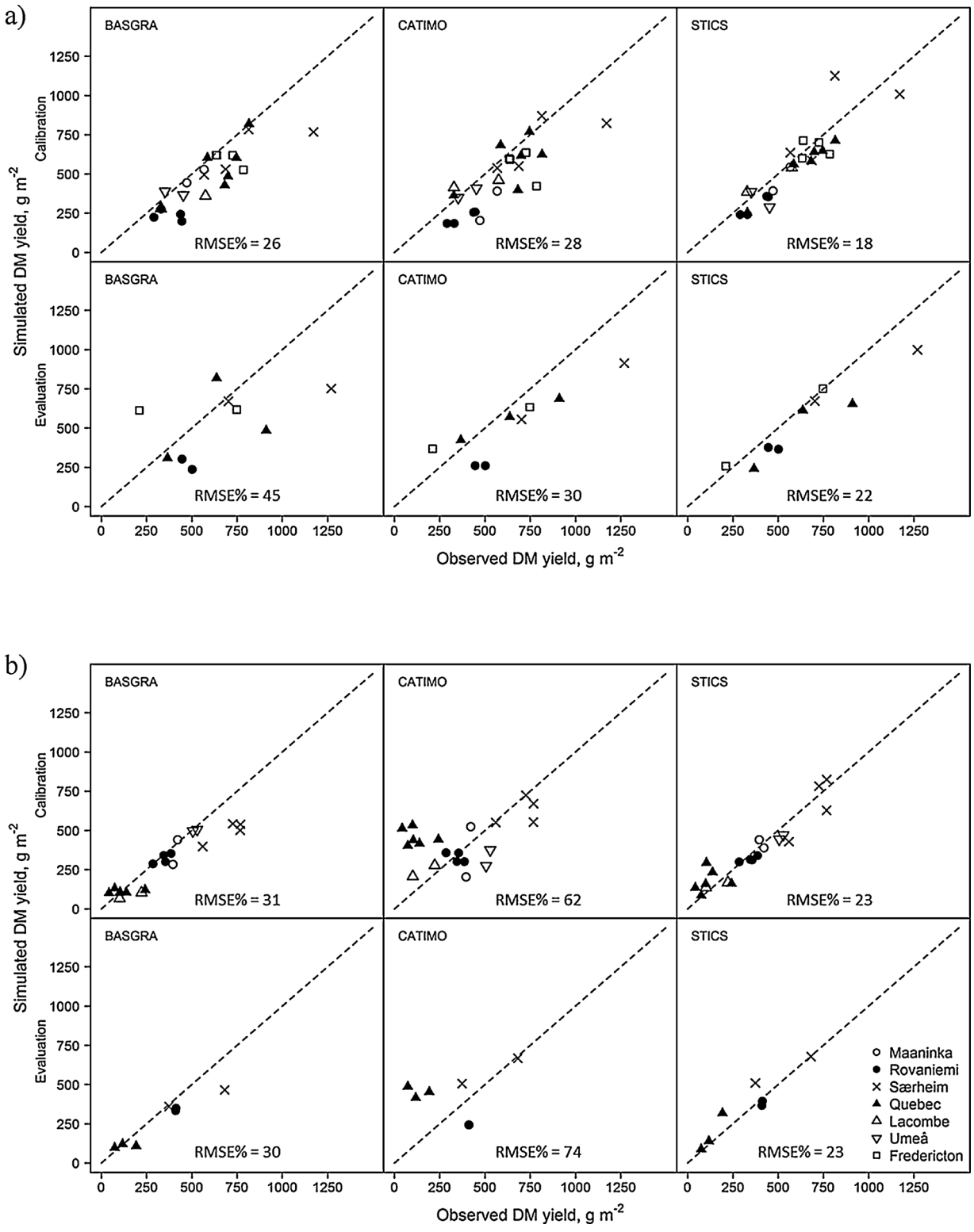

Fig. 4. Simulated (cultivar-specific calibration) and observed yield estimates of the three timothy models for: a) first cut, and b) second cut for the 24 calibrations and nine evaluation treatments. Different study sites are depicted with different symbols. The 1:1 line represents perfect agreement along with the normalized root mean square errors (RMSE%).

23% and 74% for the second cut (Fig. 4). In a previous study, Jégo et al. (2013) found the STICS model simulated timothy above-ground biomass at Canadian study sites used for model evaluation to an RMSE% of about 16% for spring growth and 89% for summer growth. The reduction of the values for the optimum and maximum temperature for photosynthesis (*teoptbis* and *temax*) compared to the previous calibration probably contributed to the better simulation of the summer growth in the STICS model.

The CATIMO model, on the other hand, has been found to simulate the yield with an accuracy of 24% (RMSE%) for the first cut and 30% for the second cut (Jing et al., 2012), which are lower figures than those observed in this study. Summer growth dynamics and the second cuts

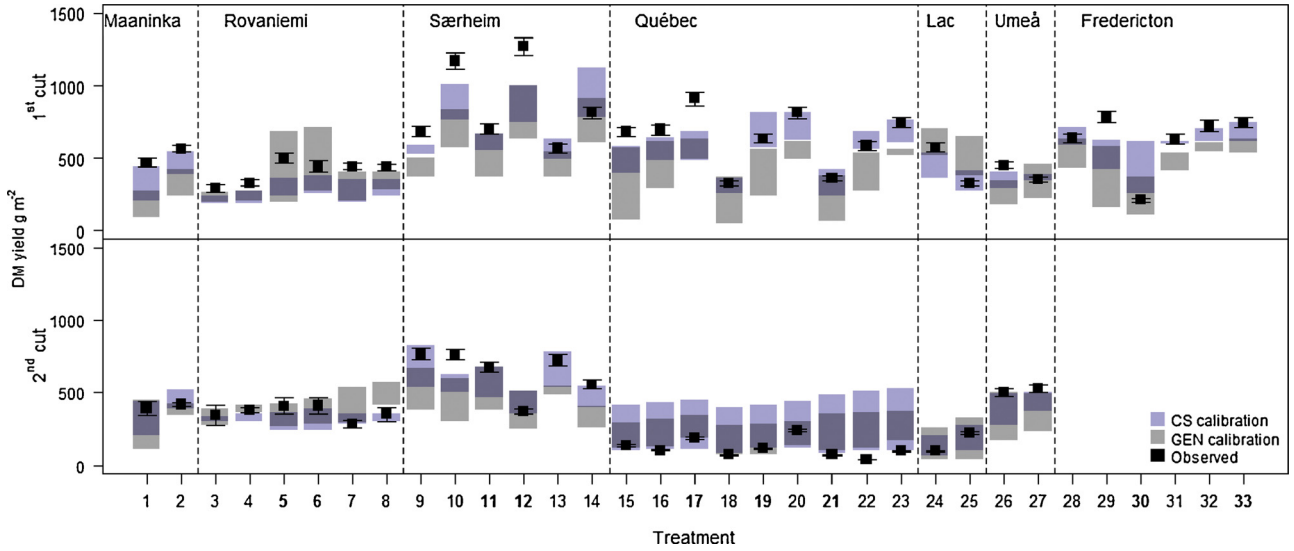

Fig. 5. Yield ranges of observations and simulations by the three models for the first cut (above) and second cut (below) for all 33 treatments at the seven study sites. The evaluation treatments are marked in bold numbers on the x-axis. Observations are marked with black squares with the standard deviation as error bars. The simulation results with cultivar-specific (CS) and generic (GEN) calibrations are denoted by blue and grey bars, respectively. The site abbreviation Lac denotes the Lacombe study site.

yields were generally well covered in the CATIMO model, but it clearly overestimated the second cut yields in Québec. At that site, there were no N applications for summer growth and the crop growth was probably N-limited. The CATIMO model could not satisfactorily take this into account. The poor accuracy of the CATIMO model in simulating the summer growth, primarily with no applied N, at the Québec site has also been noted by Jing et al. (2012) and also the poor performance for the spring growth with no applied N has been noted by Bonesmo et al. (2005) and Jing et al. (2011).

In an earlier study by Persson et al. (2014), the accuracy (RMSE%) of the LINGRA model, the predecessor of BASGRA, was similar for the first cut yield (43%) and higher for the second cut yield (46%) than was observed in this study. The BASGRA model underestimated the first cuts yields, which is in contrast with the overestimation of the first cut yield by the LINGRA model in a study by Persson et al. (2014) where the model was calibrated for Særheim in Norway and validated for several

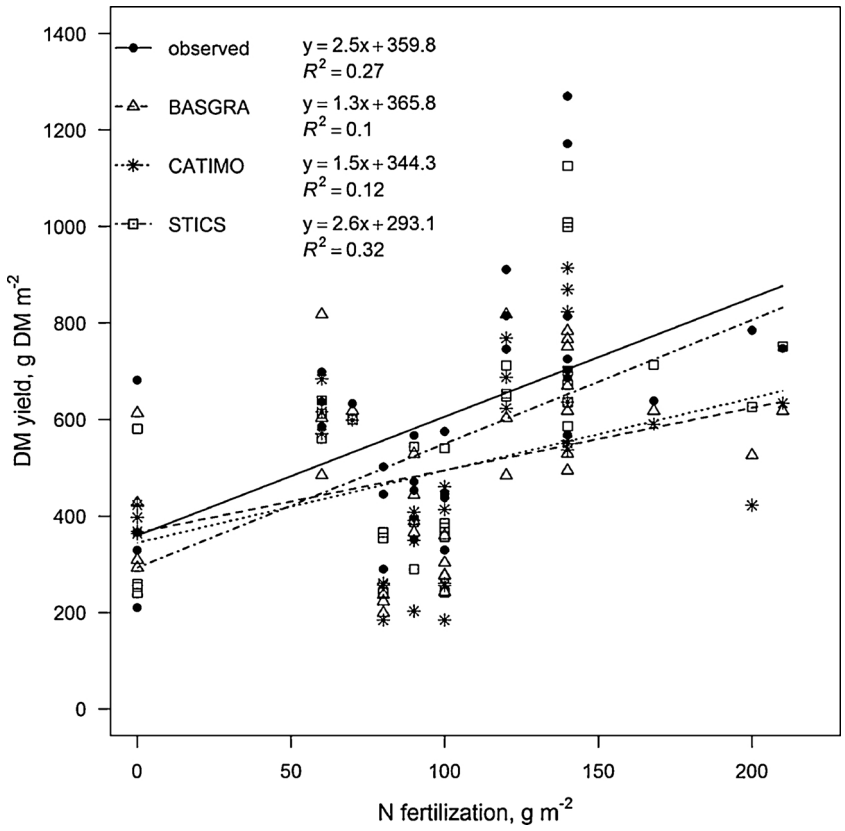

Fig. 6. Regression lines for observed and simulated first cut yields as a function of the N fertilisation rate in the spring.

locations in the Nordic region of Europe. However, the generally higher observed yields used for calibration at Særheim than at the evaluation sites might explain some of the overestimation of the yields at the latter sites. Moreover, the earlier version of the LINGRA model did not simulate nitrogen functions, which may explain some of the differences between the two model evaluations.

The simulation accuracy for the timothy models tested here was somewhat lower than that observed for the cereal models in model intercomparison studies in Europe. Reported RMSE% of grain yield estimates in those studies for winter wheat were 23–38 (Palosuo et al., 2011) and for spring barley 19–33 (Rötter et al., 2012). This is not surprising considering the challenges of modelling multi-year timothy swards related, for example, to the status of the crop after the previous growing season(s) and over-wintering, multiple cuttings and regrowth processes. Accurate simulation of crop carbon and nitrogen reserves plays an important role both for multi-annual simulations and regrowth. Additionally, the establishment of the root system and root turnover, both of which are still poorly understood processes, are more important for perennial crops than for annual crops.

Precisely simulated start of growth could be of high importance for projections of soil water status and biomass development. It has particular importance in studies projecting the impacts of climate change, as the start of growing seasons is projected to change due to increasing temperatures (Ruosteenoja et al., 2016). Here, we found differences in the simulated start of the growth of up to 55 days. However, although the STICS model simulated the start of the spring growth earlier than BASGRA and CATIMO, the early biomass accumulation was simulated relatively similarly in all models Further model evaluations where more variable winter and spring precipitation, and soil water patterns are included could reveal information about how general this insensitivity of biomass prediction to spring initialization date is.

4.2. Model calibrations and input data

In contrast to the blind-test approach applied in some earlier crop model intercomparison studies (e.g. Palosuo et al., 2011; Rötter et al., 2012), here a detailed calibration of the models was allowed. Calibration was considered particularly important as the study involved sites and cultivars from two different continents including a large variation in conditions, such as the length of the growing season and the winter period, air temperature and precipitation patterns, day length and soil type. The evaluation treatments from sites with more than two treatments were used as an independent test of the model performance. Our hypothesis that cultivar-specific calibration would work better than generic calibration was confirmed for the first cut, but for the second cut the results were not so clear. This indicates that taking into account the genetic variability between cultivars could improve estimations of the production potential of this species, particularly for the first cut. Persson et al. (2014) showed that site-specific calibration of the BASGRA model provided more precise results on the site level, whereas for the whole Nordic region, wider calibration provided better results. In our case, for some sites, the small quantity of observed data for calibration reduced the information available for cultivar-specific calibration so much that the generic calibration provided more stable results. For example, for Québec there was only one DM observation for the summer growth period.

The calibration data used in our study were exceptionally comprehensive when compared to the calibration data used in recent crop model intercomparisons (e.g., Asseng et al., 2013). The data included several sequential data series for various treatments covering altogether more than 1500 observations. The uncertainty of individual yield observations was assessed from replicates from Finnish sites (Maaninka and Rovaniemi) from where they were available. The standard deviation of those observations (27 g m^{-2}) was relatively close to that reported in the literature for yields in field trials (e.g. 10.5–14.8 g m^{-2} in Gustavsson and Martinsson (2001)).

We did not apply any weighting of the data even in the generic calibration, which means that sites with more observations had more weight than sites with fewer treatments and observations. In our case this meant, for example, that the Saerheim site in Norway got more weight in generic calibrations than other sites. However, a site by site analysis showed that no correlation was observed between model performance with generic calibration and the amount of observed data from the site used for calibration (Table S5). This result indicates that the simple approach used without weighting was reasonable.

The underestimation of the first cut yield and slight overestimation of the second cut yield may indicate that the growth descriptions of the models and the parameters related to them are too stiff when calibrated simultaneously using data from both the spring growth and the summer period. This might reflect a need for improved understanding of the processes determining timothy growth during different stages of the growing season. While using the current models for seasons with two or more cuts, it can also be hypothesized that calibrating the models with data including third and fourth cuts could lead to even higher compromise between the cuts.

4.3. Perspectives on model improvements

The STICS model gave the most accurate overall yield estimates over all treatments and this suggests that a detailed description of soil processes improves yield simulations over variable sites and growing seasons. STICS also better captured the effects of different N fertilisation rates on the first cut yield (Fig. 6), although the CATIMO model uses the same approach for simulating N effects. This result could be a consequence of a better simulation of soil N mineralisation and water balance in STICS. Both of these processes affect the N availability for plants. Generally, the simulation of N cycling, including residual N and mineralization, as well as N availability remains challenging for simulating the growth of perennial forage grasses (Ehrhardt et al., 2018).

Our analysis only concentrated on the spring growth and the first summer growth. More complete annual yield estimates would also require simulations of a second summer growth on sites and years where more than two cuts are taken. This is particularly of relevance as the growing seasons in northern latitudes are expected to get longer with changing climate (Ruosteenoja et al., 2016; Thivierge et al., 2016). The proportion of leaves to stems is usually very high in the autumn when a third cut would be taken and the nutritive value seems to differ from that of the previous cuts.

Persson et al. (2014) noticed in their LINGRA model evaluation that the model performance also depended on the ley year, since the establishment method affected the DM accumulation in the first post-seeding ley year and these effects were not covered by the evaluated model. We observed a similar trend. The first cut yield simulation of all models improved with increasing ley years at the Québec site (Table S5). This indicates that processes affecting yield development over the life cycle of timothy swards are not yet fully covered by the evaluated models.

Multi-year simulations for conditions where the winter conditions often play a major role in yield formation require models able to simulate winter survival. The data used in this study were for hardy winter cultivars in the given regions and therefore winter survival was expected to play only a minor role in yield development. For example, the winter damage sustained by the cultivar Tammisto II in northernmost location of the study, Rovaniemi, was observed to be less than 6.5% in the official variety trials (Laine et al., 2016). Although the effect of winter conditions on plant density is not simulated in STICS, Jing et al. (2017) showed that the model performed well in predicting harvested biomass, soil moisture and soil mineral nitrogen of a timothy field for three consecutive years in Eastern Canada. Of the models used in this study, the BASGRA model is able to predict the winter survival of timothy grass, but the winter module was not utilised in this study.

In addition to the yields, which were the focus of this study,

nutritive value of forage crops is of importance for an integrated assessment of the effect of climate change on farming systems. Our next step will be to compare the performance of the three models for simulating the nutritive attributes of timothy grass.

5. Conclusions

A comparison of simulation results for timothy growth in this study, using comprehensive data and advanced calibration methods, revealed substantial differences in the responses of the compared models to interannual climate and agro-management. Cultivar-specific calibration that was mostly done at site level yielded more accurate estimates of the first cut yields and preceding growth dynamics than the generic calibration applying data from all sites. Still, there was a high degree of variation among the simulated yield results and during some seasons at some sites, all three models systematically under- or overestimated the growth. This suggests that the current process descriptions in the models were not flexible enough to allow for the observed range of growth dynamics and related variables. There are obviously still needs for improvement in understanding the processes and their model descriptions, particularly regarding the summer growth and responses to nitrogen applications. These improvements are necessary to achieve reliable model applications across a wide range of current and future conditions.

Acknowledgements

Many thanks go to Vern Baron who kindly provided the data of the Lacombe site for the study. This work was conducted in the context of MACSUR (Modelling European Agriculture with Climate Change for Food Security), a project launched by the Joint Research Programming Initiative (JPI) on Agriculture, Food Security and Climate Change. The authors acknowledge financial support from the following sources: the European Agricultural Fund for Rural Development (NuRa); the Ministry of Agriculture and Forestry (FACCE-MACSUR, MalliNurmi); the Academy of Finland (NORFASYS (decision nos. 268277 and 292944), PLUMES (decision nos. 277403 and 292836) projects)); A-base funding of Agriculture and Agri-Food Canada; The Research Council of Norway and the Natural Environment Research Council of the UK through the FACCE-JPI MACSUR knowledge hub.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.fcr.2018.04.014.

References

Asseng, S., Ewert, F., Rosenzweig, C., Jones Hatfield, J.W., Ruane, J.L., Boote, A.C., Thorburn, K.J., Rötter, P.J., Cammarano, R.P., Brisson, D., Basso, N., Martre, B., Aggarwal, P., Angulo, P.K., Bertuzzi, C., Biernath, P., Challinor, C., Doltra, A.J., Gayler, J., Goldberg, S., Grant, R., Heng, R., Hooker, L., Hunt, J., Ingwersen, L.A., Izaurralde, J., Kersebaum, R.C., Müller, K.C., Naresh Kumar, C., Nendel, S., O'Leary, C., Olesen, G., Osborne, J.E., Palosuo, T.M., Priesack, T., Ripoche, E., Semenov, D., Shcherback, M.A., Steduto, I., Stöckle, P., Stratonovich, C., Streck, P., Supit, T., Tao, I., Travasso, F., Waha, M., Wallach, K., White, D., Williams, J.W., Wolf, J.R., 2013. Uncertainty in simulating wheat yields under climate change. Nat. Clim. Change 3 (9), 827–832.

Asseng, S., Ewert, F., Martre, P., Rötter, R.P., Lobell, D.B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M.P., Alderman, P.D., Prasad, P.V.V., Aggarwal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A.J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L.A., Izaurralde, R.C., Jabloun, M., Jones, C.D., Kersebaum, K.C., Koehler, A.K., Müller, C., Naresh Kumar, S., Nendel, C., O'Leary, G., Olesen, J.E., Palosuo, T., Priesack, E., Eyshi Rezaei, E., Ruane, A.C., Semenov, M.A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., Zhu, Y., 2015. Rising temperatures reduce global wheat production. Nat. Clim. Change. 5, 143–147.

Bélanger, G., Richards, J.E., 1997. Growth analysis of timothy grown with varying N nutrition. Can. J. Plant Sci. 77 (3), 373–380.

Bélanger, G., Tremblay, G.F., Mongrain, D., 2008. Yield and nutritive value of the spring growth of an ageing timothy sward. Can. J. Plant Sci. 88 (3), 457–464.

Bassu, S., et al., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Change Biol. 20 (7), 2301–2320.

Bonesmo, H., Bélanger, G., 2002a. Timothy yield and nutritive value by the CATIMO model. I. Growth and nitrogen. Agron. J. 94 (2), 337–345.

Bonesmo, H., Bélanger, G., 2002b. Timothy yield and nutritive value by the CATIMO model. II. Digestibility and fiber. Agron. J. 94 (2), 345–350.

Bonesmo, H., Bélanger, G., Charmley, E., Drapeau, R., McKenzie, D.B., Michaud, R., Tremblay, G.F., 2005. Timothy yield and nutritive value by the CATIMO model. III. validation for eastern Canada. Agron. J. 97 (1), 32–40.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.-M., Meynard, J.M., Delécolle, R., 1998. STICS: A generic model for the simulation of crops and their water and nitrogen balances: I. Theory and parameterization applied to wheat and corn. Agronomie 18, 311–346.

Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2008. Conceptual Basis, Formalisations and Parameterization of the STICS Crop Model Éditions QUAE. Versailles, France.

Chenu, K., Porter, J.R., Martre, P., Basso, B., Chapman, S.C., Ewert, F., Bindi, M., Asseng, S., 2017. Contribution of crop models to adaptation in wheat. Trends Plant Sci. 22, 472–490.

Ehrhardt, F., Soussana, J.-F., Bellocchi, G., Grace, P., McAuliffe, R., Recous, S., Sándor, R., Smith, P., Snow, V., De Antoni Migliorati, M., Basso, B., Bhatia, A., Brilli, L., Doltra, J., Dorich, C.D., Sciences, N., Doro, L., Fitton, N., Sciences, E., Giacomini, S., Grant, B., Harrison, M., Jones, S., Kirschbaum, M., Klumpp, K., Laville, P., leonard, J., Liebig, M., Lieffering, M., Martin, R., Massad, R.S., Meier, E., Merbold, L., Moore, A., Myrgiotis, V., Newton, P., Pattey, E., Rolinski, S., Sharp, J., Smith, W., Wu, L., Zhang, Q., 2018. Assessing uncertainties in crop and pasture ensemble model simulations of productivity and N2O emissions. Glob. Change Biol. 24, 603–616.

Gustavsson, A.-M., Martinsson, K., 2001. Analysis of growth and nutrition value in timothy using a dynamic model. Agr. Forest. Meteorol. 107 (2), 83–101.

Höglind, M., Schapendonk, A., Van Oijen, M., 2001. Timothy growth in Scandinavia: combining quantitative information and simulation modelling. New Phytol. 151 (2), 355–367.

Höglind, M., Hanslin, H.M., Van Oijen, M., 2005. Timothy regrowth, tillering and leaf area dynamics following spring harvest at two growth stages. Field Crops Res. 93 (1), 51–63.

Höglind, M., Van Oijen, M., Cameron, D., Persson, T., 2016. Process-based simulation of growth and overwintering of grassland using the BASGRA model. Ecol. Model. 335, 1–15.

Hurtado-Uria, C., Hennessy, D., Shalloo, L., Schulte, R.P.O., Delaby, L., O'Connor, D., 2013. Evaluation of three grass growth models to predict grass growth in Ireland. J. Agric. Sci. 151, 91–104.

Jégo, G., Bélanger, G., Tremblay, G.F., Jing, Q., Baron, V.S., 2013. Calibration and performance evaluation of the STICS crop model for simulating timothy growth and nutritive value. Field Crops Res. 151, 65–77.

Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., 2011. Modelling the biomass and harvest index dynamics of timothy. Agron. J. 103, 1397–1404.

Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P., Young, D., 2012. Regrowth simulation of the perennial grass timothy. Ecol. Model. 232, 64–77.

Jing, Q., Bélanger, G., Baron, V., Bonesmo, H., Virkajärvi, P., 2013. Simulating the nutritive value of timothy summer regrowth. Agron. J. 105 (3), 563–572.

Jing, Q., Jégo, G., Bélanger, G., Chantigny, M.H., Rochette, P., 2017. Simulation of water and nitrogen balances in a perennial forage system using the STICS model. Field Crops Res. 201, 10–18.

Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg Snow, K.V., Dimes, J.P., Silburn, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairn, D.M., Smith, C.J., 2003. An overview of APSIM: a model designed for farming systems simulation. Eur. J. Agron. 18, 267–288.

Kipling, R.P., Bannink, A., Bellocchi, G., Dalgaard, T., Fox, N.J., Hutchings, N.J., Kjeldsen, C., Lacetera, N., Sinabell, F., Topp, C.F., Van Oijen, M., Virkajärvi, P., Scollan, N., 2016. Modeling European ruminant production systems: facing the challenges of climate change. Agric. Syst. 147, 24–37.

Laine, A., Högnäsbacka, M., Kujala, M., Niskanen, M., Jauhiainen, L., Nikander, H., 2016. Results of the Official Variety Trials 2008-2015. (In Finnish) Luonnonvarakeskus (Luke). Luonnonvara- ja biotalouden tutkimus 3/2016. 239p. http://jukuri.luke.fi/bitstream/handle/10024/531720/luke-luobio_3_2016.pdf.

Nelder, J.A., Mead, R., 1965. A simplex method for function minimization. Comput. J. 7 (4), 308–313.

Nissinen, O., Kalliainen, P., Jauhiainen, L., 2010. Development of yield and nutritive value of timothy in primary growth and regrowth in northern growing conditions. Agr. Food Sci. 19, 252–268.

Palosuo, T., Kersebaum, K.C., Angulo, C., Hlavinka, P., Moriondo, M., Olesen, J.E., Patil, R.H., Ruget, F., Rumbaur, C., Takac, J., Trnka, M., Bindi, M., Caldag, B., Ewert, F., Ferrise, R., Mirschel, W., Saylan, L., Siska, B., Rötter, R., 2011. Simulation of winter wheat yield and its variability in different climates of Europe: a comparison of eight crop growth models. Eur. J. Agron. 35, 103–114.

Persson, T., Höglind, M., Gustavsson, A., Halling, M., Jauhiainen, L., Niemeläinen, O., Thorvaldsson, G., Virkajärvi, P., 2014. Evaluation of the LINGRA timothy model under Nordic conditions. Field Crops Res. 161, 87–97.

Pirttioja, N., Carter, T.R., Fronzek, S., Bindi, M., Hoffmann, H., Palosuo, T., Ruiz-Ramos, M., Tao, F., Trnka, M., Acutis, M., Asseng, S., Baranowski, P., Basso, B., Bodin, P., Buis, S., Cammarano, D., Deligios, P., Destain, M.F., Dumont, B., Ewert, F., Ferrise, R., François, L., Gaiser, T., Hlavinka, P., Jacquemin, I., Kersebaum, K.C., Kollas, C., Krzyszczak, J., Lorite, I.J., Minet, J., Minguez, M.I., Montesino, M., Moriondo, M.,

Müller, C., Nendel, C., Öztürk, I., Perego, A., Rodríguez, A., Ruane, A.C., Ruget, F., Sanna, M., Semenov, M.A., Slawinski, C., Stratonovitch, P., Supit, I., Waha, K., Wang, E., Wu, L., Zhao, Z., Rötter, R.P., 2015. Temperature and precipitation effects on wheat yield across a European transect: a crop model ensemble analysis using impact response surfaces. Clim. Res. 65, 87–105.

Rötter, R.P., Palosuo, T., Kersebaum, K.C., Angulo, C., Bindi, M., Ewert, F., Ferrise, R., Hlavinka, P., Moriondo, M., Nendel, C., Olesen, J.E., Patil, R.H., Ruget, F., Takác, J., Trnka, M., 2012. Simulation of spring barley yield in different climatic zones of Northern and Central Europe: a comparison of nine crop models. Field Crops Res. 133, 23–36.

Ruiz-Ramos, M., Ferrise, R., Rodríguez, A., Lorite, I.J., Bindi, M., Carter, T.R., Fronzek, S., Palosuo, T., Pirttioja, N., Baranowski, P., Buis, B., Cammarano, D., Chen, Y., Dumont, B., Ewert, F., Gaiser, T., Hlavinka, P., Hoffmann, H., Höhn, J.G., Jurecka, F., Kersebaum, K.C., Krzyszczak, J., Lana, M., Mechiche-Alami, A., Minet, J., Montesino, M., Nendel, C., Porter, J.R., Ruget, F., Semenov, M.A., Steinmetz, Z., Stratonovitch, P., Supit, I., Tao, F., Trnka, M., De Wit, A., Rötter, R.P., 2017. Adaptation response surfaces for managing wheat under perturbed climate and CO 2 in a Mediterranean environment. Agr. Syst 159, 260–274.

Ruosteenoja, K., Räisänen, J., Venäläinen, A., Kämäräinen, M., 2016. Projections for the duration and degree days of the thermal growing season in Europe derived from CMIP5 model output. Int. J. Climatol 36 (8), 2029–2055.

Thivierge, M.N., Jégo, G., Bélanger, G., Bertrand, A., Tremblay, G.F., Rotz, C.A., Qian, B., 2016. Predicted yield and nutritive value of an alfalfa?timothy mixture under climate change and elevated atmospheric carbon dioxide. Agron. J. 108 (2), 585–603.

van Oijen, M., Höglind, M., Hanslin, H.M., Caldwell, N., 2005. Process-based modeling of timothy regrowth. Agron. J. 97 (5), 1295–1303.

Virkajärvi, P., Pakarinen, K., Hyrkäs, M., Savolainen, J., Isolahti, M., Schnyder, H., Schellberg, J., 2010. Does tiller type distribution explain the differences in yield and nutritive value of timothy genotypes. Grassland Sci. Eur. 15, 572–574.

Virkajärvi, P., Hyrkäs, M., Pakarinen, K., Rinne, M., 2012. Timotein ja ruokonadan erot sadontuottoprosessissa. In: Hyrkäs, M., Virkajärvi, P. (Eds.), Nurmen kasvu- ja kehitysprosessit. MTT Report 56 (In Finnish), pp. 22–46.